INCREASING COMPUTATIONAL EFFICIENCY FOR SEMI-SEPARABLE MARKOV DECISION PROCESSES

Roy Mendelssohn
Southwest Fisheries Center
National Marine Fisheries Service, NOAA
Honolulu, Hawaii 96812

January 1978
1st revision February 1978
2d revision March 1978

DRAFT FOR COMMENT

I. Introduction

One of the major deterrents to using Markov decision processes (MDP) in real applications is the extremely large problem size that may arise. Several authors (Koehler, Whinston, and Wright [4]; Puterman [5]; and references cited by these papers) have attacked this problem by developing more efficient solution algorithms for a general MDP. In this paper, the more standard linear programming (LP) approach is used. For a wide class of MDP's that arise in real applications, the resulting LP can be greatly reduced in dimensionality. This allows for efficient aggregation of the LP into smaller, easily solved problems, as well as easily derived qualitative properties of solutions.

The class of problems considered is similar to the separable programs considered by Denardo [1]. However, the method of proof is entirely different. This allows for the relaxation of the assumption that the one period return function be separable. The final rows of the reduced LP can be interpreted in terms of the distribution function of leaving each state; this fact, and the readily obtained qualitative properties of a solution can often lead to even greater reduction in problem size.

II. The Model and Main Results

Problems that arise in the context of capital accumulation and consumption, managing renewable resources, reservoir management, and inventory control and production, involve solving the following mathematical program:

$$f(x) = \text{maximum } \{G(x, y) + \alpha Ef(s[y, D]) : y \in Y(x)\} \qquad x \in X \tag{2.1}$$

where X is the set of states, Y(x) is the set of feasible decisions for each $x \in X$, and D is a random variable. These are called "semi-separable" problems (c.f., Denardo [1]) since the transition function depends only on the decision, not on the state.

It is well known (see d'Epenoux [2]) that an optimal policy to the discrete problem, if one exists, is a solution to the following LP:

$$\min \sum_{x \in X} f_x \tag{2.2}$$

$$\text{subject to } \sum_{j \in X} (\delta_{xj} - \alpha P_j^y) f_j \geq G(x, y) \quad y \in Y(x),\ x \in X$$

$$\text{where } \delta_{xj} = \begin{cases} 0 & \text{if } x \neq j \\ 1 & \text{if } x = j \end{cases}$$

and P_j^y is the discrete probability of going to state j given decision j. P_j^y can be obtained by discretizing s[y, D]. Fox [3] gives a discussion of methods of discretizing a continuous problem. Note also that the usual problem has P_{xj}^y. However, by the semi-separable assumption, $P_{xj}^y = P_j^y$.

The equivalent dual problem is:

$$\max \sum_{x \in X} \sum_{y \in Y(x)} G_x^y v_x^y$$

$$\text{subject to:} \quad \sum_{x \in X} \sum_{y \in Y(x)} (\delta_{xj} - \alpha P_j^y) v_x^y = 1 \quad , \quad j \in X \qquad (2.3)$$

$$v_x^y \geq 0 \quad \text{for all } y \in Y(x);\ x \in X$$

where G_x^y is the discrete equivalent of $G(x, y)$.

In this section, attention is restricted to x, y scalar. Section III shows that this is without loss of generality. $Y(x)$ is assumed to be of the form:

$$Y(x) = \{y : 0 \leq y \leq x\} \ . \qquad (2.4)$$

The cardinality of X, $|X|$, is assumed to be n+1.

Theorem 2.1 shows that semi-separable MDP's have the property that it is possible a priori to reduce the MDP to a MDP that has two actions per state.

Theorem 2.1 For each j, $0 \leq j \leq n$, let $b_j^* = \underset{0 \leq i \leq j}{\text{maximum}} \left\{G_j^i - G_{j-1}^i\right\}$, and let i_j^* be defined as the i where b_j^* obtains its maximum. Let $A(j)$ be an optimal policy function. Then:

$$\text{either } A(j) = j$$

$$\text{or } A(j) = i_j^*$$

Proof: Consider the dual LP (2.3). Since the rows are equalities, add rows 1 through n to row 0 (row 0 is the first row constraint, row n the last row constraint), rows 2 through n to row 1, etc., yielding a new LP:

$$\text{maximize} \sum_{x=0}^{n} \sum_{i=0}^{x} G_x^i v_x^i \tag{2.5}$$

$$\text{s.t.} \sum_{i=0}^{n} \sum_{x=i}^{n} \left\{ \delta_{x \geq j} - \alpha \sum_{k=j}^{n} P_k^i \right\} v_x^i = n+1-j \qquad j = 0, 1, \ldots, n$$

$$v_x^i \geq 0 \qquad x = 0, \ldots, n; \quad i = 0, 1, \ldots, x$$

where $$\delta_{x \geq j} = \begin{cases} 0 & x < j \\ 1 & x \geq j \end{cases}$$

Transform variables as follows: for each i, i = 0, 1, ..., n

$$w_n^i = v_n^i$$
$$w_j^i - w_{j+1}^i = v_j^i \qquad j = i, i+1, \ldots, n \tag{2.6}$$

or equivalently:

$$\begin{aligned} w_n^i &= v_n^i \\ w_{n-1}^i &= v_n^i + v_{n-1}^i \\ &\vdots \\ w_i^i &= v_n^i + v_{n-1}^i + \ldots + v_i^i \end{aligned} \tag{2.7}$$

The LP now becomes:

$$\text{maximize} \sum_{x=0}^{n} \sum_{i=0}^{x-1} \left(G_x^i - G_{x-1}^i\right) w_x^i + \sum_{x=0}^{n} G_x^x w_x^x$$

$$\text{s.t.} \sum_{i=0}^{n} \left\{\delta_{x \geq j} - \alpha \sum_{k=j}^{n} P_k^i\right\} w_i^i + \sum_{\substack{i=0 \\ i \neq j}}^{j} w_j^i = n+1-j \qquad j = 0, 1, \ldots, n \tag{2.8}$$

$$w_n^j \geq 0 \qquad j = 0, 1, \ldots, n$$

$$w_j^i - w_{j+1}^i \geq 0 \qquad j = 0, 1, \ldots, n-1, \; i = 0, 1, \ldots, j$$

For an arbitrary j, $1 \leq j \leq n$, and for some i, $0 \leq i_j < j$, $i_j \neq i_j^*$, transform variables by:

$$\bar{w}_j^{i_j} = w_j^{i_j}$$

$$\bar{w}_j^{i_j^*} = w_j^{i_j} + w_j^{i_j^*}$$

or equivalently:

$$\bar{w}_j^{i_j} = w_j^{i_j}$$

$$\bar{w}_j^{i_j^*} - \bar{w}_j^{i_j} = w_j^{i_j^*} \tag{2.9}$$

Then the LP (2.8) becomes:

$$\text{maximize} \sum_{x=0}^{n} \sum_{\substack{i=0 \\ i \neq i_x^* \\ i \neq i_j}}^{n} \left(G_x^i - G_{x-1}^i\right) w_x^i + \sum_{x=0}^{n} G_x^x w_x^x$$

$$+ \sum_{\substack{x=1 \\ x \neq j}}^{n} \left(G_x^{i_x^*} - G_{x-1}^{i_x^*}\right) w_x^{i_x^*} + \left(G_j^{i_j^*} - G_{j-1}^{i_j^*}\right) \bar{w}_j^{i_j^*} + \left(\left(G_j^{i_j} - G_{j-1}^{i_j}\right) - \left(G_j^{i_j^*} - G_{j-1}^{i_j^*}\right)\right) \bar{w}_j^{i_j}$$

$$\text{s.t.} \sum_{i=0}^{n} \left\{ \delta_{x \geq y} - \alpha \sum_{k=y}^{n} P_k^i \right\} w_i^i + \sum_{i=0}^{n} w_y^i = n+1-y \qquad \begin{matrix} y = 0, 1, \ldots, n \\ y \neq j \end{matrix}$$

$$\sum_{i=0}^{n} \left\{ \delta_{x \geq j} - \alpha \sum_{k=j}^{n} P_k^i \right\} w_i^i + \sum_{\substack{i=0 \\ i \neq i_j^* \\ i \neq i_j}}^{n} w_j^i + \bar{w}_j^{i_j^*} = n+1-j \qquad (2.10a)$$

$$w_n^i \geq 0 \qquad i = 0, 1, \ldots, n$$

$$w_y^i - w_{y+1}^i \geq 0 \qquad \begin{matrix} y = 0, 1, \ldots, n-1; \\ i = 0, 1, \ldots, y \end{matrix}$$

$$w_{j-1}^{i_j} - \bar{w}_j^{i_j} \geq 0 \qquad (2.10b)$$

$$w_{j-1}^{i_j^*} - \bar{w}_j^{i_j^*} \geq -\bar{w}_j^{i_j}$$

$$\bar{w}_j^{i_j} - w_{j+1}^{i_j} \geq 0$$

$$\bar{w}_j^{i_j^*} - w_{j+1}^{i_j^*} \geq \bar{w}_j^{i_j}$$

$$w_j^{i_j^*} - \bar{w}_j^{i_j} \geq 0$$

Examine the constraints involving $\bar{w}_j^{i_j}$:

$$\bar{w}_j^{i_j} \geq \bar{w}_j^{i_j^*} - w_{j-1}^{i_j^*}$$

$$\bar{w}_j^{i_j} \geq w_{j+1}^{i_j}$$

$$\bar{w}_j^{i_j^*} \geq \bar{w}_j^{i_j}$$

$$w_{j-1}^{i_j} \geq \bar{w}_j^{i_j}$$

$$\bar{w}_j^{i_j^*} - w_{j+1}^{i_j^*} \geq \bar{w}_j^{i_j}$$

Since $\bar{w}_j^{i_j}$ only appears in the constraints (2.10b), and has a nonpositive objective function coefficient, at an optimal solution it will have a value at a lower bound, which is the greater of $w_{j+1}^{i_j}$, $w_j^{i_j^*} - w_{j-1}^{i_j^*}$.

If $\bar{w}_j^{i_j} = w_{j+1}^{i_j}$, then from (2.7) and (2.9), $v_j^{i_j} \equiv 0$.

If $\bar{w}_j^{i_j^*} - w_{j-1}^{i_j^*} > w_{j-1}^{i_j}$, then from (2.) $w_j^{i_j} = \left(w_j^{i_j^*} - w_j^{i_j}\right) - w_{j-1}^{i_j^*}$

which implies:

$$2w_j^{i_j} = w_j^{i_j^*} - w_{j-1}^{i_j^*}$$

From (2.8),

$$0 \geq w_j^{i_j^*} - w_{j-1}^{i_j^*} = 2w_j^{i_j} \geq 0$$

so that $w_j^{i_j} \equiv 0$. Since $w_j^{i_j} \geq w_{j+1}^{i_j}$, then $w_j^{i_j} = w_{j+1}^{i_j} = 0$, which implies $v_j^{i_j} \equiv 0$. Since j and i_j were chosen arbitrarily, the proof holds for each j and i_j.

□

The transformations in theorem 2.1 are only necessary to prove that an optimal policy chooses either x or i_x^*. Corollary 2.1 gives the reduced LP.

Corollary 2.1 In (2.2), let i_j^* be defined as in theorem 2.1. Then, an optimal solution to:

$$\begin{aligned} \text{minimize} \quad & \sum_{x=0}^{n} f_x \\ \text{s.t.} \quad & \sum_{x=0}^{n} \left(\delta_{ix} - \alpha P_x^i\right) f_x \geq G_i^i \qquad i = 0, 1, \ldots, n \\ & \sum_{x=0}^{n} \left(\delta_{i_j^* x} - \alpha P_x^{i_j^*}\right) f_x \geq G_j^{i_j^*} \qquad j = 0, 1, \ldots, n \end{aligned} \tag{2.11}$$

is also an optimal solution to (2.2).

Proof: Immediate consequence of theorem 2.1.

□

An alternate form to (2.11), which is sparser when computations are done is:

$$\text{minimize} \sum_{x=0}^{n} f_x$$

$$\text{s.t.} \quad \sum_{x=0}^{n} \left(\delta_{ix} - \alpha P_x^i\right) f_x - \lambda_i^i = G_i^i \qquad i = 0, 1, \ldots, n \tag{2.12}$$

$$f_j - f_{i_j^*}^* + \lambda_{i_j^*}^{i_j^*} - \lambda_j^{i_j^*} = G_j^{i_j^*} \qquad j = 0, 1, \ldots, n$$

$$\lambda_i^i \,,\ \lambda_j^{i_j^*} \geq 0 \qquad \begin{matrix} i = 0, 1, \ldots, n \\ j = 0, 1, \ldots, n \end{matrix}$$

which is derived by making the constraints in (2.11) equalities, and then subtracting the appropriate rows.

For certain special cases, the results of theorem 2.1 can be made stronger, and derived more directly. This is especially true of separable MDP's (Denardo [1]). Let $G(x, y) = a(x) + b(y)$. Transform variables by:

$$f_0 = u_o$$
$$f_i = \sum_{j=0} u_i \qquad i = 1, \ldots, n$$

or equivalently:

$$u_i = f_i - f_{i-1} \qquad i = 0, 1, \ldots, n$$

It is straightforward to show that at an optimal solution:

$$f_i^* \geq \left(a(i) - a(i-1)\right) + f_{i-1}^* \tag{2.13}$$

when equation (2.13) is valid, all the rows equivalent to choosing j as the action are redundant with:

$$\sum_{x=0}^{n} \left(\delta_{xy} - \alpha P_x^j\right) f_x \geq a(j) + b(j)$$

$$f_{j+1} - f_{j+i-1} \geq a(j+i) - a(j+i-1)$$

$$i = 1, 2, \ldots, n-j$$

This leads to a reduced LP:

$$\text{minimize} \quad \sum_{x=0}^{n} (n+1-x)u_x$$

$$\text{s.t.} \quad \sum_{x=0}^{n} \left(\delta_{x\leq j} - \alpha \sum_{i=x}^{n} P_i^j\right) u_x \geq a(j) + b(j) \quad j = 0, 1, \ldots, n \tag{2.14}$$

$$u_x \geq a(x) - a(x-1) \quad x = 1, 2, \ldots, n$$

where $\delta_{x\leq j}$ denotes:

$$\delta_{x\leq j} = \begin{cases} 1 & x \leq j \\ 0 & x > j \end{cases}$$

The LP in (2.14) can be solved with a $(n+1 \times n+1)$ constraint matrix. The equivalent formulation in Denardo (1968) has a $(2n+2 \times n+1)$ constraint matrix.

III. Extensions and Discussion

The proof of theorem 2.1 does not depend on x, y being scalars. It only depends on the semi-separable assumption, and the assumption the $Y(x) = \{y : 0 \leq y \leq x\}$. The vector case is proven in a similar manner. The dual variables are again grouped by choosing action y from all states $x \geq y$, and then making the same transformations of variables.

The LP (2.11) can be solved using linear programming, or else can be solved using successive approximations or similar iterative techniques (see Koehler et al. [4]; Puterman [5]) with a reduced action space. The reduced LP can also be used to prove qualitative results about optimal policies when further assumptions are made on the form of $G(\cdot, \cdot)$. This will be explored more fully in a future paper.

Even with the reduction in problem size between (2.11) and (2.2), for many real applications any useful discrete state space will still bring about an exceedingly large LP. However, the reduced LP gives more insight into how to effectively aggregate rows or columns in order to further reduce the size of the LP, using results similar to those in Zipkin [6].

REFERENCES

[1] Denardo, E. V., "Separable Markovian Decision Problems," Management Science 14(1968):451-462.

[2] d'Epenoux, F., "Sur un Problème de Production et de Stockage dans l'Aléstoire," Revue Française de Recherche Opérationnelle 4(14)(1960), p. 3.

[3] Fox, B. L., "Discretizing Dynamic Programs," Journal of Optimization Theory and Applications 11(1973):228-234.

[4] Koehler, G. J., A. B. Whinston, and G. P. Wright, "Optimization over Leontief Substitution systems," North-Yolland/American Elsevier, N.U.(1974), 221 p.

[5] Puterman, M., "Modified Policy Iteration Methods of Dynamic Programming," Paper presented at Joint National ORSA/TIM meeting, Nov. 7-9, 1977, Atlanta, Georgia.

[6] Zipkin, P. H.,"A Priori Bounds for Aggregated Linear Programs with Fixed-Weight Disaggregation. Technical Report #86, School of Organization and Management, Yale University (1977), 37 p.